The U.S. Census Bureau is in the preliminary fact-finding stages of an Integrated Information Solution project, of which DADS will be a part. The project defines the details of how the U.S. Census Bureau should proceed from DADS development to the full vision of an integrated information access and dissemination system for internal and external users. We envision that DADS will play a major part in the U.S. Census Bureau's new world of "Electronic Commerce."

The American FactFinder subsystem is accessible through the U.S. Census Bureau's Intranet and the World Wide Web via a standard browser. It allows users to select census products or submit queries against varied datasets to extract meaningful information and produce results, which can be in table or map form. Some queries may be generated based upon maps and most results can be reformatted, including formatting from tables-to-maps or vice versa.

The Data Product Production subsystem will provide an interactive interface for designing, reviewing and generating 1998 Dress Rehearsal and Census 2000 data products, including Public Law 94-171 (PL) listings and summary files, the Sample Summary File, and the Hundred Percent Summary File. In addition, the Data Products Production subsystem will, based on these data products, generate a number of file out-put formats for printing, for stamping to CD-ROM, and for online Intranet/ Internet access through the American FactFinder.

1.1 DADS Products, Services, and Customers

As previously mentioned, The American FactFinder is a corporate, enterprise-wide subsystem designed to provide access to, creation of, and delivery of census data products and data sets. The American FactFinder allows customers to select census products (existing on-line documents) or query against data sets (collections of data that can be correlated/tabulated). Executed queries generate results that can be in table or map form and can also, for the most part, be reformatted, including reformatting from tables-to-maps or vice versa. Internal government staffs may eventually use the American FactFinder in response to expressed customer needs to produce new products by assembling sets of queries and appropriately formatting the results. Once approved, these products will become standard U.S. Census Bureau products and will then be made available to other customers through the American FactFinder.

Analysis of current trends indicate that future customers will need to be able to generate queries across data in new and unique ways, correlating information not envisioned by a single survey but as part of diverse surveys. In response to this need, the American FactFinder will incorporate the capability for customers to associate and integrate data from different datasets. Eventually, we envision that non-census datasets will become accessible through the American FactFinder, even though they are not hosted by it.

Data Product Production Data Products	
Class	**Description**
1998 Dress Rehearsal Public Law 94-171 (PL)	• Public Law listing for printing; • ASCII file of metadata/data for the CD-ROM product and for FTP download from the Internet; and • Internal format for upload into the DADS98 Inquiry system for access to Tier 2 products
Hundred Percent Summary File (HSF)	• ASCII file of metadata/data for the CD-ROM product; and • Internal format for upload into the DADS98 inquiry system for access to Tier 2 products
Sample Summary File (SSF)	• ASCII file of metadata/data for the CD-ROM product; and • Internal format for upload into the DADS98 inquiry system for access to Tier 2 products

The following table lists some of our clients and the data products they disseminate through DADS.

Data Access and Dissemination System Clients: U.S. Census Bureau		
Name	**Census or Survey Name**	**Census/Survey Description**
Decennial Directorate	• 1990 Decennial Census; • 1998 Dress Rehearsal for Census 2000; and • 2000 Decennial Census	Census of population and housing taken every 10 years to determine how many Congressional representatives each state will have
Economic Directorate	• Economic Census	Conducted in years ending in "2" and "7," the Economic Censuses compile information about businesses, industries, establishments, products, and local areas
Demographic Directorate	• American Community Survey • Continuous Measurement	A new survey being developed by the U.S. Census Bureau to provide, every year, economic, social, and housing information about communities and population groups

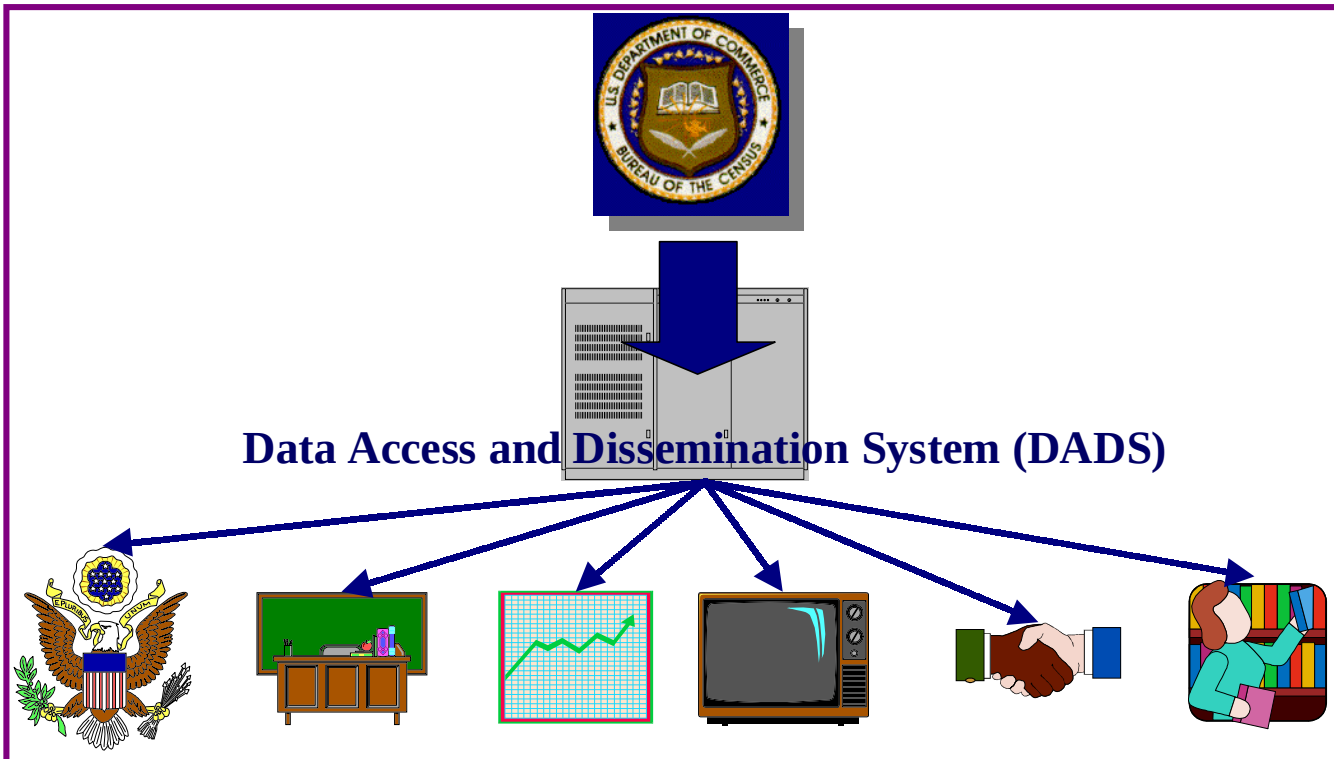

Figure 1: Data Access and Dissemination System (DADS) Customers

2.0 DADS IT Support

DADS is designed to support many U.S. Census Bureau staffs and organizations, and as such is a resident member of the corporate network infrastructure, relying upon support from the Telecommunications Office at Headquarters.

Figure 2, below, depicts how the American FactFinder subsystem supports IT.

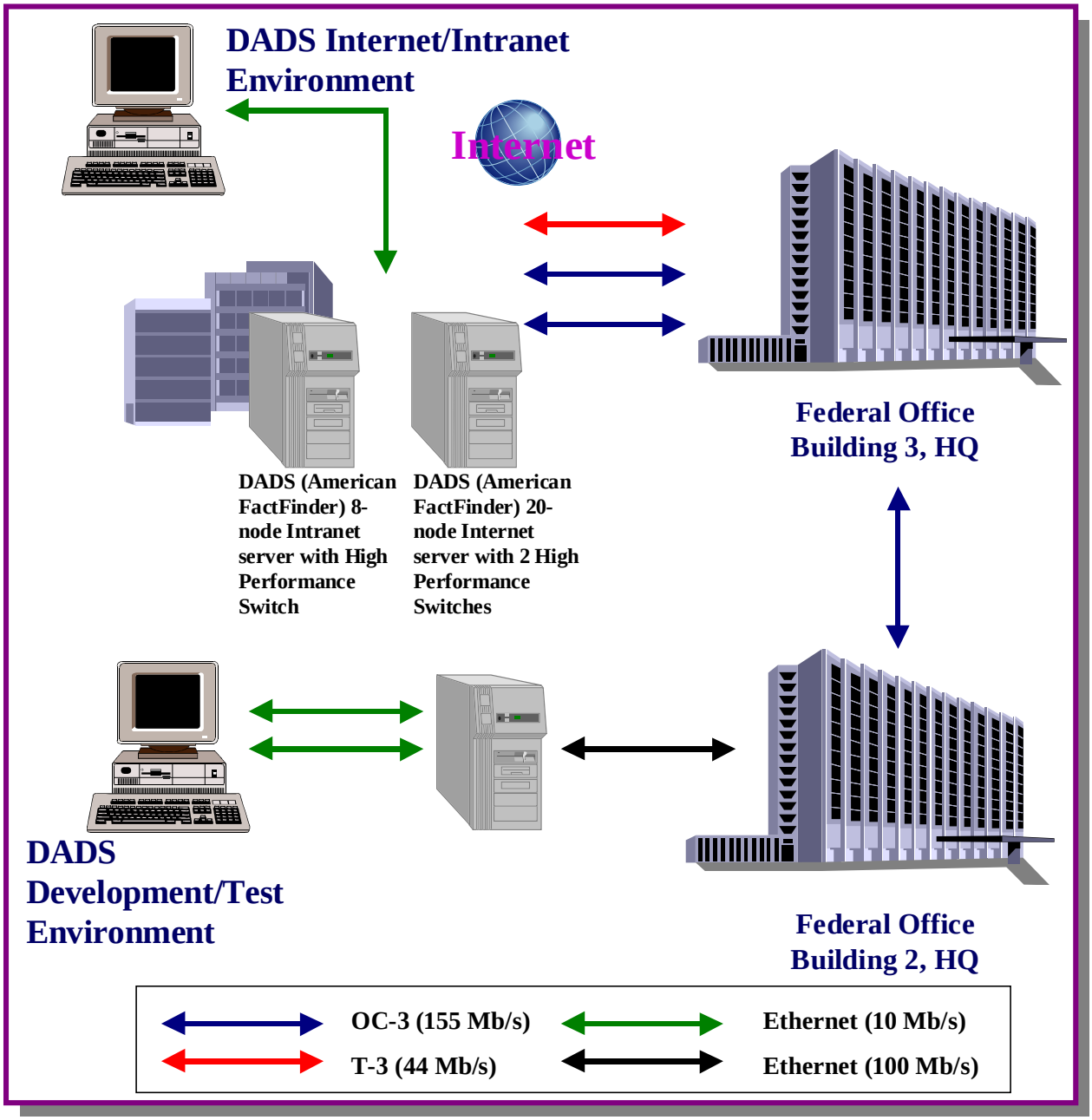

Figure 2: American FactFinder IT Support

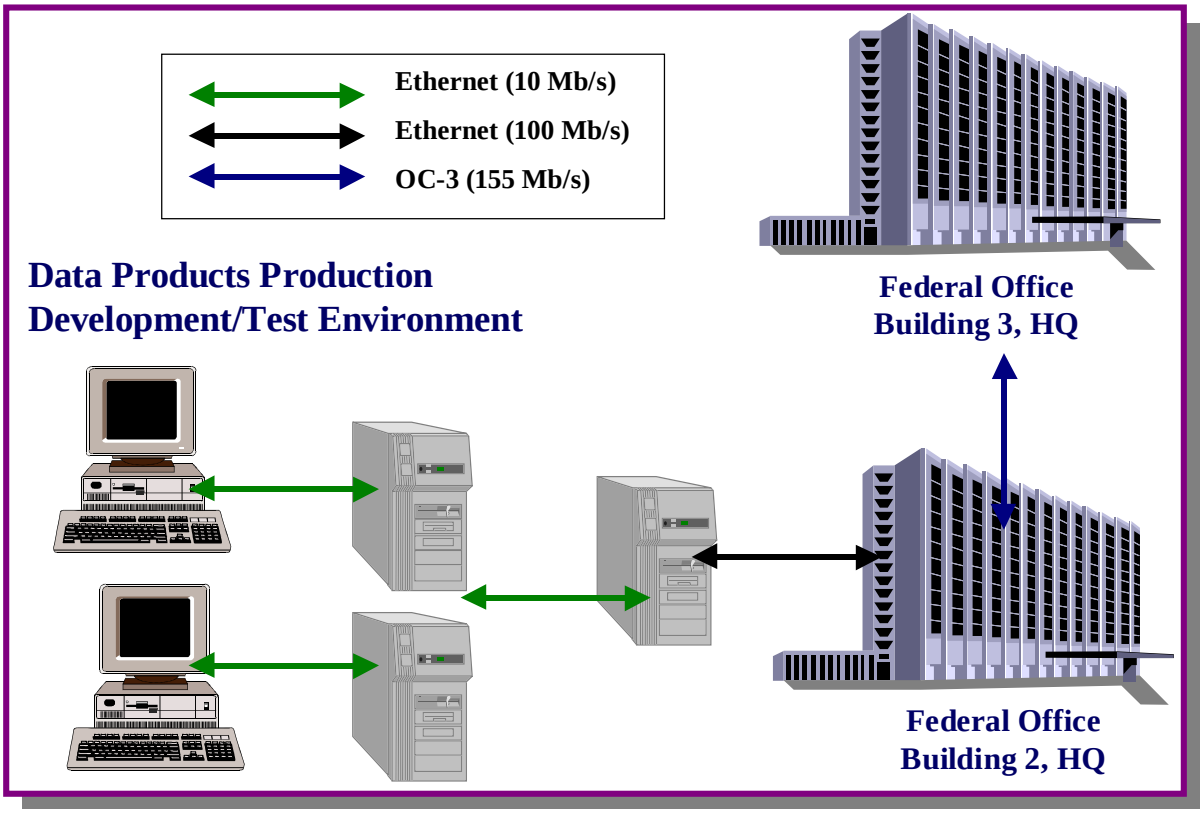

Figure 4: Data Products Production IT Support

2.1 DADS IT System Description

2.1.1 Detailed Description of American FactFinder

The IBM RS/6000 SP family of servers acts as host to the American FactFinder subsystem and is located at the Bowie Computer Center. This family of servers provides the widest possible range of scalability for servers of its class (Unix and NT), allowing for scalability and extendibility into the future. Because of its clustered configuration, the SP isolates most software and hardware failures, providing, overall, a very high level of reliability. American FactFinder software is open in nature and allows development and deployment on other, similar hardware. Internal and external servers are separate due to U.S. Census Bureau security and confidentiality protection concerns.

Each of the SP servers consists of several Symmetric Multi Processing shared-memory nodes. In the initial configuration, each node consists of four high-performance PowerPC processors and 3GB of memory, along with I/O interfaces. Other nodes are available which have different attributes. All nodes are connected with a high-performance switch whose aggregate bandwidth is proportional to the number of connections between nodes. Software is installed on nodes such that no single node failure can take the system down.

The American FactFinder configuration is shown in Figure 5, on the following page.

The following table lists the American FactFinder subsystem's major hardware components.

American FactFinder (AFF) Major Hardware Components							
Quantity	**Vendor/ Model**	**CPU**	**RAM**	**Storage**	**Com-ponents or Modules**	**Physical Location**	**Com-ments**
2	IBM RS/6000 F40P Work-stations	333 MHz	256 Mb	9 GB	HWACS	BCC	100MB/s NIC
2	IBM RS/6000 43P Work-stations	333 MHz	128 Mb	18 GB	150 MB Switch	HQ	100MB/s NIC
1	IBM RS/6000 SP Server w/8 nodes	4 per node @333 MHz	3 GB/node	1 TB	High Speed Switch, SMP, 3570 Tape	BCC	100MB/s NIC per node
1	IBM RS/6000 SP Server w/20 nodes	4 per node @333 MHz	3 GB/node	1 TB	(2) High Speed Switches, SMP, ADSM	BCC	100MB/s NIC per node

The manner in which all software packages are installed on the RS/6000 SP ensures that scalability and reliability is maximized. Each of the significant software packages is installed on multiple nodes, ensuring operation even if a node fails, while multiple nodes also provide additional processing power. No single node failure can compromise system integrity and the RS/6000 SP nodes have been divided according to software demands (based on memory needs, CPU needs, I/O throughput, etc.). The following table lists the American FactFinder subsystem's major software components.

American FactFinder (AFF) Major Software Components	
Software	**Description**
IBM AIX and components	IBM's AIX is the primary DADS RS6000/SP operating system. It is a 64-bit, reliable, scalable variant of the UNIX operating system, geared for parallel processing in symmetric and distributed parallel environments. It includes components such as the High Availability Cluster Monitor Package and High Availability Control Workstation providing reliable operation, even under degraded conditions. The DADS AIX installation includes a comprehensive backup facility (ADSM) which manages a tape library and powerful monitoring and reporting tools.
ESRI: SDE and IMS	The ESRI COTS family of software products satisfies the majority of our geographic and mapping needs. Although some customization of the COTS code is necessary, we have fully utilized the COTS ESRI Internet Map Server (IMS) and Spatial Data Engine (SDE).
Oracle Appli-cation Server	Oracle's OAS is key to the DADS AFF architecture. It provides Web service and a framework for custom server-side application software in a manageable environment that spans hardware. OAS is optimized for deployment across multiple nodes (or separate computers), and handles much of the administrative burden of balancing processing load.

It is our intent to build upon the DADS foundation to create the Integrated Information Solutions (IIS) capability. The IIS program consists of many capabilities that integrate our geographic and census/survey data (both numeric and demographic) and that move us toward more effective and universal electronic data dissemination. The items listed below for the American FactFinder subsystem form a significant component of our Census Modernization initiative.

American FactFinder New Requirements	
Requirement	**Description**
Redundant Sites	**State Data Centers, and other governmental and large-agency data organizations, require highly responsive, high-capacity support for their data operations. Redundant DADS sites would duplicate the most resource-intensive aspects of DADS and update the data at these sites on a real-time basis.**
Downloading Files	**Many high-level users and organizations are clamoring for the ability to access and transfer very large Census data files (very large Economic PDF files, for example) for off-line review and analysis on third-party applications. It is most efficient and cost-effective to establish a DADS FTP site, that can be managed as part of the overall DADS data management process, that maintains desired files and provides them for download.**
Electronic Commerce	**The U.S. Census Bureau has mandated that some types of data available through the external DADS AFF subsystem will have associated charges. In order to support such a requirement, the AFF subsystem will need to be able to monitor all queries originating from the user, identify those that have charges associated with them, collect all such charges, and conduct a purchase transaction.**
Batch Processing	**The DADS Team anticipates that users will produce very long and complex queries. These queries, if implemented in real time by the system, would consume enormous resources, reduce the overall responsiveness of DADS, and make less of the system available for other users. This requirement would provide a means for the system to process a query during non-peak conditions.**
Home Page Integration	**Census end users will get confused if there is a proliferation of styles, menu choices, icons, etc., when they access the U.S. Census Bureau web site. Consolidation of web designs around the current DADS user interface design would greatly reduce confusion and create an integrated, consistent look and feel for all of census data.**
Enhancements for Economic Data	**The Economic Directorate has extremely demanding requirements for dissemination of their data. It is essential that significant attention be paid to enhancing DADS' capabilities to completely display Economic data, otherwise vital information will be lost and studies based on that information will be compromised.**
Performance Enhancements	**Census data products have never before been made available via the Web in the manner envisioned for AFF. It is anticipated that, because of the innovative and user-friendly interface and the depth of information available, the American FactFinder will be very popular to in-depth users as well as casual browsers. Current estimates for usage are based on U.S. Census Bureau historical data, which are considered untrustworthy since this type of approach has never been used before; engineering estimates are that workload on the AFF internal and external systems will nearly triple by 2001.**
Data Integration	**One of the U.S. Census Bureau's primary goals is to find means to intelligently and effectively integrate the data from the program areas. It is important to ensure consistency and accuracy, both for U.S. Census Bureau data and for the retrieval and dissemination systems for accessing the data.**
Integration of Geography	**This would allow DADS to provide comprehensible support to the widest user community with a consistent, integrated view of geography.**

External American FactFinder Delay Factors	
Factor	**Description**
Budget	**Redefining and clarifying existing requirements consumed more of the budget than originally planned. To accomplish necessary DPP subsystem development in FY98/99, further AFF requirements analysis was postponed. These analyses will be completed as early as possible so that design/development work on DADS2000 can commence.**
Technology	**To assure the designed level of performance, IBM accomplished a technology upgrade to the system designated for the external component of AFF. This established both external and internal systems in identical types and levels of technology and introduced the ability to load balance by moving components between individual systems as needed. This was done at no cost to the government.**
User responsiveness as a result of operational experience	**DADS is a tool rather than a process and therefore is sensitive to customer feedback and highly responsive to their communicated needs.** **Even with the extensive design modeling and theoretical design extrapolations, preliminary user feedback was somewhat attenuated by the lack of a hands-on system. Once our customers had an AFF to test drive, we began to gather truly effective data which, to an extent, influenced further design direction. Many of the suggestions and issues raised were deemed to be significant enough to warrant a delay in public release until we could act on them.**

American FactFinder Milestones, FY 98					
Description	**Estimated**		**Actual**		**Progress to Date**
	Start Date	**Finish Date**	**Start Date**	**Finish Date**	
No milestones to report					

American FactFinder Milestones, FY 99					
Description	**Estimated**		**Actual**		**Progress to Date**
	Start Date	**Finish Date**	**Start Date**	**Finish Date**	
Initial Release of Production AFF	**01/99**	**01/99**		**01/99**	**Completed.**
Security Plan approval	**01/99**	**02/99**		**02/99**	**Completed.**
Conduct program/System Risk Assessment	**02/99**	**02/99**		**02/99**	**Completed.**
Release 2 of Production AFF	**06/99**	**06/99**		**06/99**	**Completed.**
Security Certification and Accreditation	**06/99**	**06/99**	**01/99**	**02/99**	**Temporary, valid for six months.**
Security Certification and Accreditation (Final)	**08/99**	**08/99**			

American FactFinder Milestones, FY 03					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Security Plan Review and Revision	02/03	03/03			
Production Operation	10/02	09/03			
AFF System Enhancement/ Requirements Gathering, Review and Analysis	10/02	09/03			
Post Production Enhancement	10/02	09/03			

2.1.3 American FactFinder Performance Measures

American FactFinder Performance Measures			
Performance Goals	Performance Measures	Target Performance	Current Performance
Availability of the American FactFinder subsystem to act as a tool in support of production and dissemination requirements	% of goals using AFF products	100%	100%
	% of performance error rate	100%	100%
	% user friendliness and product quality	100%	100%
	% of system availability	98%	98%
	% of competence after training	98%	98%
American FactFinder subsystem capable of producing and disseminating acceptable data products on time and in the quantities required	% of goals using AFF products	100%	100%
	% of performance error rate	100%	100%
	% user friendliness and product quality	100%	100%
	% of system availability	98%	98%
American FactFinder will serve a larger customer base with greater speed, accuracy and more tailored customer service	% of goals using AFF products	100%	100%
	% of performance error rate	100%	100%
	% of system availability	98%	98%
Availability of Tier 1,2,3 data access and dissemination capabilities on required dates and in accordance with response time requirements	% of performance error rate	100%	100%
	% of capacity availability	100%	0%
System performance	% of planned capacity	100%	100%
	% of planned throughput	100%	100%
	% of planned response times	100%	100%

the Public Law 94-171 (PL) listings and summary files is mandated by law and failure to meet that schedule will incur significant legal repercussions.

Our risk mitigation strategies include the following:

- **establishing effective working relationships with supplier organizations to foster a co-operative working environment and to establish a clear understanding that timely delivery of accurate data files and products is in everyone's best interest;**
- **supporting corporate buy-in and acceptance of DADS as the primary tool for census data dissemination;**
- **setting milestones for the development process to bilaterally monitor the progress of data and data products destined for dissemination through DADS;**
- **establishing a tracking mechanism/tool to address those milestones; and**
- **establishing and publishing "drop dead" dates as management tools to meet the established schedules and projections.**

Another risk is that DADS might not meet established deadlines due to its inability to acquire and/or retain staff with requisite critical skills. DADS depends heavily upon skilled government staff with appropriate policy and practice expertise made available to this program only through the cooperation of the program areas. Additionally, because we are hindered by the General Schedule salary guidelines, DADS is not competitive on the open market when searching for the requisite technical skills. Training the staff has proven ineffective since the extensive cutting-edge technical and in-depth subject matter experience needed cannot be taught and those personnel with high-demand skills often leave for better career opportunities.

DADS faces a tangible risk in having difficulty finding qualified and effective replacements, and in losing the cumulative programmatic knowledge and experience. This has grave effects on current development efforts and future intentions involving the transfer of technology to government staff and the eventual assumption of program responsibilities by government staff. This could make DADS contractor-dependent for an extended period.

Our risk mitigation strategies include the following:

- **defining clear career paths for government and providing necessary training. We will place conditions on personnel receiving U.S. Census Bureau-sponsored training to ensure that they stay with us;**
- **effectively using the IT Services Contract for support to find the highly specialized technical skills necessary to DADS' success;**
- **effectively using the contractor support to achieve the necessary flexibility in staffing to maintain the appropriate number and types of personnel to meet workload levels and technical requirements;**
- **maximizing the salary pool to effectively compensate qualified staff;**
- **effectively transferring technology through the Integrated Product Team process. Government staff must be fully involved and cognizant of contractor efforts; and**
- **standardizing hardware and using commercial off-the-shelf software.**

Another risk is that the U.S. Census Bureau's confidentiality and security requirements might be compromised because DADS uses Internet technology to disseminate data. For any web-based application, malicious or frivolous access, i.e., hacking, is a major concern. The connection to the World Wide Web, Internet technology concepts, such as firewall security and the implementation of confidentiality filters, all

With DADS permanently housed at the Bowie Computer Center, we must rely on the resident organization to address contingency operational issues. We have serious concerns about the lack of a current and regularly tested Site Contingency Plan that addresses the particularities and requirements of the IBM SP environment.

Our risk mitigation strategies include the following:

- **working with the U.S. Census Bureau and Bowie Computer Center staffs to develop a comprehensive U.S. Census Bureau Contingency Plan to activate in the event of anomalous conditions at the Bowie Computer Center;**
- **beginning contingency planning focused on addressing the need for a redundant, mirrored, or "hot" site;**
- **establishing a reciprocal agreement(s) with existing sites to serve as backup to DADS in the event of system failures, catastrophic events, or acts of vandalism; and**
- **securing funding to implement a duplicate system or capability at a second, geographically separate site.**

2.1.5 American FactFinder References

The American FactFinder subsystem is supported by the following planning documents:

- **Budget Submission for FY 2001, dated June 1999;**
- **Census 2000 Systems Architecture Report, dated March 1, 1999;**
- **Census Information Technology Review Guide, dated February 2, 1999;**
- **DADS98 AFF Release 1 Test Results, dated January 4, 1999;**
- **1999 Strategic IT Plan, dated December 18, 1998, pages 28, 61, 63, 93;**
- **DADS Y2K Compliance Review, dated October 10, 1998;**
- **DADS Program Management Plan, continuous updates from June 1997 to February 1999;**
- **DADS Comprehensive Development Plan, continuous updates from June 1997 to February 1999;**
- **Requirements Analysis Documentation for DADS Inquiry System, continuous updates from June 1997 to February 1999;**
- **DADS Architecture Plan (version 2.0), continuous updates from June 1997 to January 1999;**
- **DADS Business Case Analysis Report, May 1997;**
- **DADS 1996 Test Results, dated March 17, 1997;**
- **1997 DADS Analysis Report, dated March 11, 1997; and**
- **Census Security Plan CEN230, continuous updates from February 1996 to February 1999.**

The following table lists the Data Products Production subsystem's major software components.

Data Products Production Major Software Components	
Software	**Description**
Microsoft Windows NT	**Microsoft's Windows NT is the operating system running on the Compaq servers that support the DPP Component of DADS. Windows NT is a 32-bit, reliable operating system geared for single- and small-parallel processing in symmetric environments. It includes monitoring and reporting tools, and is integrated with Novell software to allow access to remote file systems.**
SuperCROSS	**SuperCROSS is a COTS integrated data aggregation and dissemination system software technology that performs tabulations across large datasets with minimal consumption of computer system resources. We are using it specifically for tabulating Decennial large-scale statistical data. One of the main components is SuperSTAR, which supports client/server architecture and currently runs on Windows 95/NT systems. STR is porting the SuperSTAR suite to AIX and DADS will move all existing SuperCROSS processing to the IBM RS6000/SP by late 2000.**
SAS	**We use SAS for transforming files from one format to another as part of the DPP process. We use SAS programs to process incoming data (from data providers) into a SuperCROSS compatible format, and extract information from SuperCROSS into other necessary forms.**
HTML, URL, HTTP	**Communication between the client and the server is done strictly with web protocols: HTTP network traffic, carrying URLs from the client and HTML (version 3.2) from the server. This protocol is a widely accepted worldwide standard and is typically passed through firewalls. DADS uses this API for communicating between the internal and external servers.**

SuperCROSS is purpose-built for Data Products Production-type tabulations and is used at other national statistical agencies. It offers:

- **a graphical interface well suited to component and tabular specification;**
- **native understanding of important concepts such as dimensions, universes, recodes, and geographical hierarchies;**
- **the ability to handle hierarchically organized input data; and**
- **a production module capable of volume production of statistical tables.**

SuperCROSS operates with Windows 95/NT clients running against a local database or against a Windows NT server.

It is our intent to build upon the DADS foundation to create the Integrated Information Solutions (IIS) capability. The IIS program consists of many capabilities that integrate our geographic and census/survey data (both numeric and demographic) and that move us toward more effective and universal electronic data dissemination. The items listed on the following page for the Data Products Production subsystem form a significant component of our Census Modernization initiative.

Data Products Production Milestones, FY 99					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Conduct DPP System Risk Assessment	02/99	02/99		02/99	Completed.
Conditional Acceptance of NT based DPP system	02/99	02/99		02/99	Completed.
Security Certification and Accreditation	06/99	06/99	01/99	02/99	Temporary, valid for six months.
Initial release of PL 94-171 data	04/99	04/99		04/99	Completed.
Formal Acceptance of NT based DPP system	04/99	07/99			In progress.
Security Certification and Accreditation (Final)	08/99	08/99			
Port to AIX/SP Platform	09/99	09/99			

Data Products Production Milestones, FY 00					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Y2K Compliance Testing	04/99	10/99			In progress.
Y2K Compliance Certification	10/99	10/99			
Requirements Gathering	12/98	12/99			In progress.
Y2K Full Regression Testing for Verification	11/99	12/99			
Requirements Analysis, Planning and Documentation	01/99	01/00			In progress.
Security Plan Review and Revision (DPP Component)	02/00	03/00			
AIX-based Development	09/99	07/00			
Production Operation	04/99	09/00	01/99		
Integration Testing	07/00	09/00			

Data Products Production Milestones, FY 01					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Training	09/00	10/00			
Acceptance Testing	09/00	10/00			
Deployment for Census 2000	10/00	11/00			
Security Plan Review and Revision	02/01	03/01			
Production	12/00	09/01			
DPP System Enhancement/ Requirements Gathering, Review and Analysis	12/00	09/01			
Post Production Enhancement	12/00	09/01			

2.2.4 Data Products Production Risks

Please see section 2.1.4, American FactFinder Risks.

2.2.5 Data Products Production References

The Data Products Production subsystem is supported by the following planning documents:

- **Budget Submission for FY 2001, dated June 1999;**
- **Census 2000 Systems Architecture Report, dated March 1, 1999;**
- **Census Information Technology Review Guide, dated February 2, 1999;**
- **DADS98 AFF Release 1 Test Results, dated January 4, 1999;**
- **1999 Strategic IT Plan, dated December 18, 1998, pages 28, 61, 63, 93;**
- **DADS Y2K Compliance Review, dated October 10, 1998;**
- **DADS Program Management Plan, continuous updates from June 1997 to February 1999;**
- **DADS Comprehensive Development Plan, continuous updates from June 1997 to February 1999;**
- **Requirements Analysis Documentation for Data Products Production (DPP) System, continuous updates from June 1997 to February 1999;**
- **DADS Architecture Plan (version 2.0), continuous updates from June 1997 to January 1999;**
- **DADS Business Case Analysis Report, May 1997;**
- **DADS 1996 Test Results, dated March 17, 1997;**
- **1997 DADS Analysis Report, dated March 11, 1997;**
- **1997 DADS DPP Analysis Report; and**
- **Census Security Plan CEN230, continuous updates from February 1996 to February 1999.**

3.0 DADS Infrastructure Description

3.1 Detailed Description of DADS Infrastructure

DADS' infrastructure is supported by the Decennial program area.

3.2 DADS Infrastructure Progress Against Planned Milestones

Data Access and Dissemination Systems (DADS) Infrastructure Milestones, FY 98

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
No milestones to report					